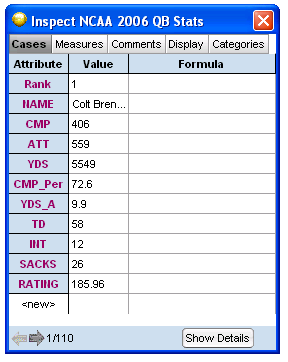
**AP Stat- Introduction to FATHOM!**

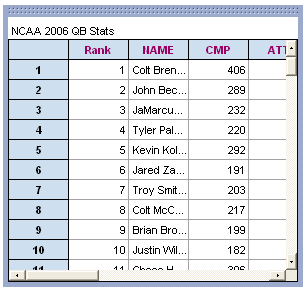
**Open Fathom 2**

* Start->Programs->Fathom 2->Fathom 2

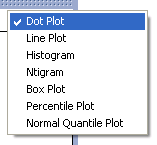
**Opening an Existing File**

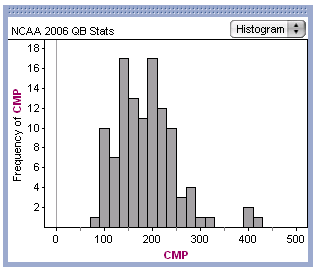
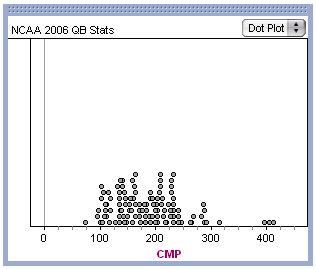
* File->Open
* Go to the “STH\_SHARES:” Drive and find the “McNelis” Folder
* Open the SHARE folder, then the STAT folder, and then Open “NCAA 2006 QB Stats”
* In the Blank Fathom Document a collection will drop in.
* Double-click on the collection
* The inspect collection window will appear with all the attributes (variables) and the values of a single observation (see picture on right)
* At the bottom LEFT you will see the case number you are in (#1) as well as how many observations are present (110)
* Flip through- Can you identify the 10th quarterback and his # of TD’s?
* Close out of the Inspection window by clicking the red X.

**Opening a spreadsheet/table**

* Click once on the collection so that it is highlighted
* Fathom uses a lot of drag-drop features
* In the menu, click and hold on the table icon
* Drag the table into the white field
* A spreadsheet should appear with all the attributes on the top of the columns and observations in the cells.
  + If the data does not appear: grab the name of the collection, drag it to the table and drop it in.
* Resize the table so you can see all the columns
* To get rid of a table simple hit delete. The data will not disappear (it is still in the collection. You just closed the table)!
* Bring back the table (recreate it in your document)

**Creating Graphs**

* In the menu click and hold on the graph icon and Drag it into the white field
* A blank graph will appear. An attribute must be dragged into the graph
* From the Table grab the attribute “CMP” and drag it to the bottom axis of the graph
* A dot plot will automatically be drawn
* Change this to a histogram by clicking on the drop-down menu in the top corner and selecting histogram. Also try changing it to a boxplot.



* Go back to a histogram. To change the bin (bar) width, double click into the white area of the graph. An “Inspect Graph” box will appear.
* Resize this box so you can see all the categories
  + “binAlignmentPosition” tells the program where to start the first bin (where to start your x-axis)
  + “binWidth” tells the program how wide the bins should be
* Change the graph so that the first bin starts at 50 and the width of each bar is 30
* Now the bars go too high- they are off the graph. Fix this by changing the “yUpper” number or **grab the top of the y-axis and drag down** (this is cooler!). Make 25 your upper bound.
* Change the attribute being graphed: grab “ATT” and drag it over “CMP” on the graph. Drop it and it should change the histogram to “ATT”
* Change the start of the first bin to 150, and the bin width to 50. Make sure to adjust the graph so you can see all the bars completely. What is the shape of this distribution?
* Select one bar in the graph. You will notice that the observations that are in that bar are now highlighted in the spreadsheet.

**Copying Graphs to Word Documents**

* To copy a graph to a word document (or to a power point), select the entire graph “ATT”
* In the menu select Edit->**Copy as Picture** or hit **Ctrl+Shift+C**
* Open a word document. In the document select Edit->Paste or hit Ctrl+V
* Once you have accomplished this, you can close the word document without saving.

**Creating New Collections & Inputting your own data**

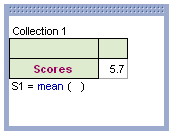
* Open a new, blank fathom document.
* Grab the Collection icon in the menu and drag it to a blank field 
* An empty box will be shown.
* Open a table for the collection (drag the table icon to the blank screen)
* Click on the **<new>** attribute. Rename it “**Scores**”
* Enter the following data:

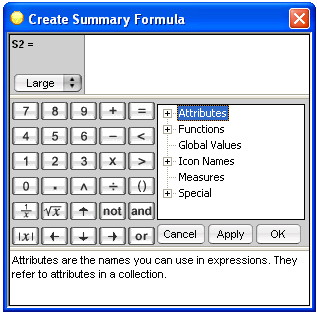
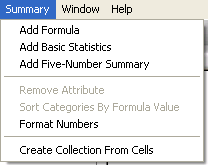
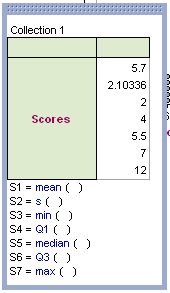
6, 8, 3, 5, 7, 2, 4, 6, 8, 4, 3, 7, 9, 5, 6, 4, 7, 6, 3, 5, 4, 5, 5, 8, 7, 6, 7, 5, 4, 12

* Create a new attribute and call it “Gender”
* Enter the following data:

M, M, M, F, F, M, F, F, M, M, M, F, F, F, F, F, M, M, M, F, F, F, F, M, M, F, F, F, M, F

* Create a histogram of SCORES.
* Create a boxplot of SCORES. Let’s say we wanted to look at the differences in scores between the two genders. Drag the attribute GENDER to the y-axis of the graph. This should create 2 parallel boxplots, one for Male and one for Female.

**Creating Summary Tables**

* Grab the Summary icon in the menu and drag it to your Fathom document
* A blank summary table will appear. Drag the “Scores” attribute to the left side of the summary table.
* The only statistic given is the mean. To add more statistics select “Summary” in the drop down menu at the top of the page. Select “Add Formula”
* To add the standard deviation, type “s()” into the formula.
* Hit OK, then resize the table so you can see the #’s
* Select the “Summary” menu again and select “Add Five-Number Summary”
* Resize the table. You can now see all vital statistics for SCORES.
* If you want to find the statistics of SCORES broken down by Gender, drag the attribute GENDER to the top of the summary table you created. This should separate the statistics by Males and Females.

Open a new Fathom document.

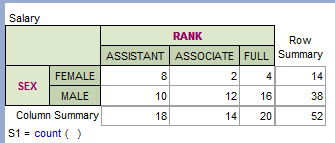
Go to File->Import->Import From File

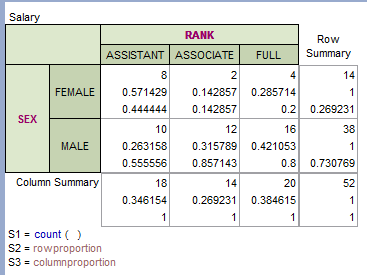
In the SHARE folder again, open Salary.TXT

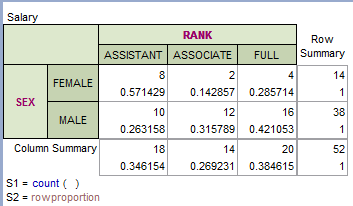
The collection should appear in the window.

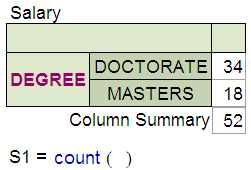
Open a spread sheet for this collection so you can see all the attributes/variables.

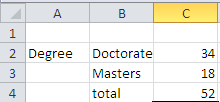
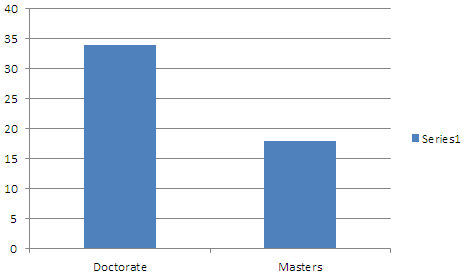
**Creating Two-Way Tables**

* Pull down a new summary table.
* On the left of the table drag the attribute “SEX”
* On the top of the table drag the attribute “RANK”
* This will create a two-way table since both attributes are categorical.
* To create the conditional distribution of SEX, select the “Summary” menu and select “Add Formula”
* In the formula box type “rowproportion” and select “OK”
* To create the conditional distribution of RANK, select the “Summary” menu and select “Add Formula”
* In the formula box type “columnproportion” and select “OK”

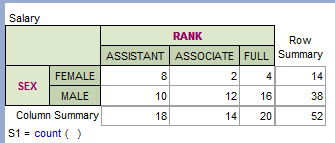
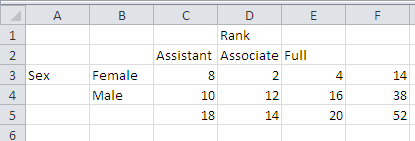
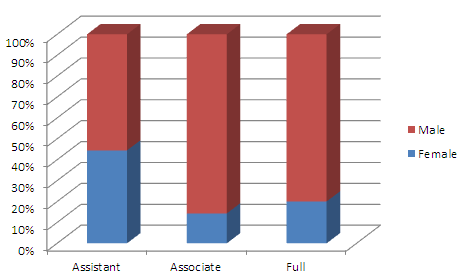


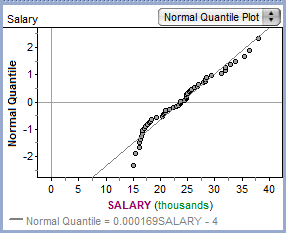
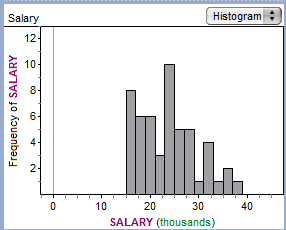


**Creating Bar Graphs (using EXCEL)**

* Open an Excel document
* Create a summary table of a categorical variable. Drag down a Summary table, and then drag a categorical variable (like DEGREE) into the table.
* Transfer the table you just created (that you created on Fathom) to Excel by hand (you cannot copy and paste)- jus type the info into excel as shown to the left.
* Highlight just the data (don’t include where is says “degree” and don’t include the total).
* Go to Insert 🡪 Column 🡪 2D column 🡪 and pick the one at the top left.
* This will create a bar chart for you. You can edit the title and other things on the chart.
* You can also select 3D column, for a fancier picture.
* Pie charts can also be selected.
* To create a stacked (segmented) bar chart, first create a 2 way table (with 2 categorical variables).

(Use the SEX vs. RANK one you created earlier)

* Transfer the table to Excel by hand
* Highlight the data (again, don’t include totals) and go to Insert 🡪 2D or 3D column 🡪 and use the 3rd one over (the one where the bars go all the way up to the top).

**Testing for Normality**

* To see if a set of data fits a Normal model, we use a Normal Quantile Plot.
* Pull down a new graph. Drag in the Attribute “SALARY” into the bottom.
* From the drop-down menu choose “Normal Quantile Plot”
* The more the data fits along the line the closer it will fit a Normal Distribution. For this data a right skew can be seen in the Normal Quantile plot from the points on the left not fitting.

**PRACTICE #1:**

1. Open a blank word document and put your name at the top of the page.
2. On the next line, put “PRACTICE #1”, then hit enter
3. In Fathom, using the NCAA QB Data, create 2 histograms: one for “YDS\_A” and another for “TD”
4. Copy these into the word document.
5. Find the full summary stats of each variable, and copy the summary tables into the word document.
6. Save the file as your: lastname.firstname (example: McNelis.Lauren) to your student folder on the computer. Leave the document open.
7. Go back to Fathom and exit out of “NCAA 2006 QB Stats” (do not exit out of fathom, just the document you were working on!)

**PRACTICE #2**

Open a new fathom document, new collection and a new table. Add the data for the following 3 variables:

RESPONSE: y, y, y, n, n, y, y, y, y, n, n, n, y, y, y, y, y, n, n, n, n, y, y, y, y, y, y

AGE: 12, 15, 16, 14, 15, 16, 17, 14, 15, 16, 14, 17, 18, 16, 15, 14, 11, 17, 12, 14, 11, 12, 14, 15, 16, 16, 17

GENDER: m, f, f, m, m, m, f, f, f, m, m, f, f, m, f, m, f, m, f, m, m, m, f, f, f, f, m

Complete the following using the data above. Copy all things into a word document as you do them. Be sure to title a new page in the word doc with “PRACTICE #2”

1. Fins the full summary statistics for AGE. Also create a boxplot & histogram of the ages.
2. Find the full summary statistics for AGE broken down by GENDER. Also, create parallel boxplots of this.
3. Create a one-way table (summary table) of RESPONSE. Then transfer to excel and create a bar chart.
4. Create a two-way table of RESPONSE vs. GENDER. Then transfer to excel and create a segmented bar chart.
5. What percent ***of males*** said “Y”?
6. Check the normality of the AGE variable. Does it seem to be normally distributed? Use the Empirical Rule (68-95-99.7).

* First, find the mean and standard deviation.
* Next, find the ranges of µ ± σ, µ ± 2σ, and µ ± 3σ.
* Then look thru your data and find the % of data in each of these ranges.
* Compare these percentages to 68-95-99.7

1. Save the word document again.

**Making Scatterplots**

Go back to the NCAA 2006 collection again

Let’s make a scatterplot of “Attempts” versus “Touchdowns”

* Drag down a new graph. Grab the attribute “ATT” and drag it to the x-axis.
* Grab the attribute “TD” and drag it to the y-axis of the graph.
* You should have a scatterplot of ATT vs. TD



**Finding correlation coefficient**

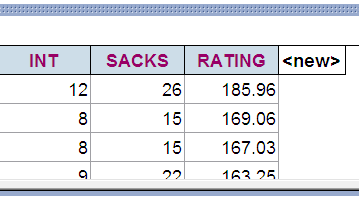
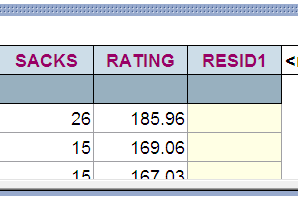
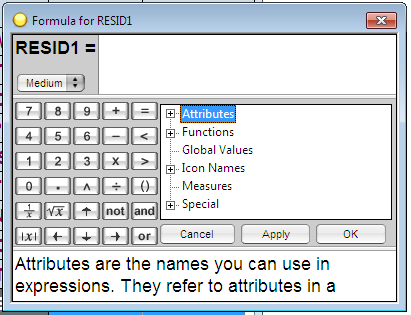
* Drag down a new summary table.
* Drag the attribute “ATT” to the left row
* Drag the attribute “TD” to the top column
* The correlation coefficient should be stated in the center.

**Finding the LSR line, and the Residual plot**

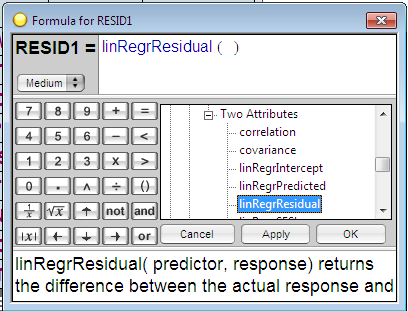
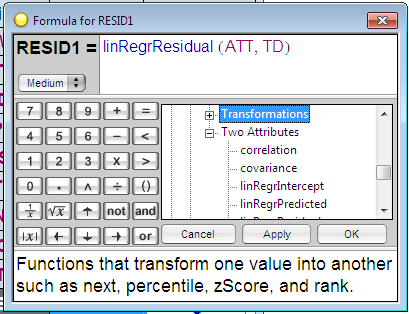
* While the scatterplot is highlighted (use the same one above with ATT vs. TD), go to the drop-down menu GRAPH and click on “Least-Squares line”
* You will notice that the LSR line has been added to your scatterplot and the equation and r2 are listed down at the bottom of the plot.
* To make the residual plot: Make sure the graph is still highlighted, and go to the menu GRAPH again, and this time click on “Make Residual Plot”
* The residual plot will appear below the scatterplot. Make the entire picture bigger so you can clearly see the residual plot.



**Another way to get the residual plot… finding the list RESID**

* We will use the same plot as above (ATT vs. TD).
* Find the collection NCAA QB Stats 2006 on your Fathom document and create a spreadsheet so you can see ALL the variables.
* Scroll across the spreadsheet until you see a spot for a new variable <NEW>
* In this spot, type in the new variable RESID1
* Go to the TABLE menu at the top of your screen, and click on SHOW FORMULAS.
* Double click on the gray area under RESID
* This will bring up a formula window
* Click on the + sign next to “Functions,” then scroll down and click the + sign next to “Statistical”
* Click the + sign next to “Two Attributes”
* Double click on “LinRegrResidual”. This

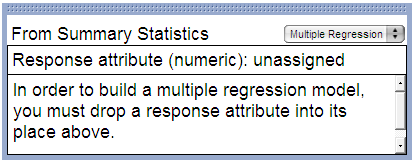
will then appear in the top box.

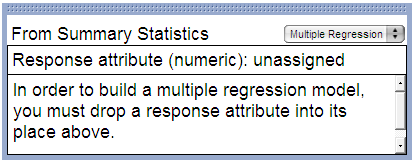
* You now need to put in (X variable, Y variable)
* To do this, scroll back up and click the + sign next to “Attributes” and then find your X variable, double click on it, type a comma, then find your Y variable, double click on it. Then hit the OK button
* Now you have the variable RESID in your spreadsheet. You can use create a residual plot by dragging down a GRAPH and putting the X variable on the X axis and the list RESID on the Y-axis
* You can also now check to see if your residuals are normally distributed by making a normal quantile plot of the list RESID. Do this.

**Linear Regression Output**

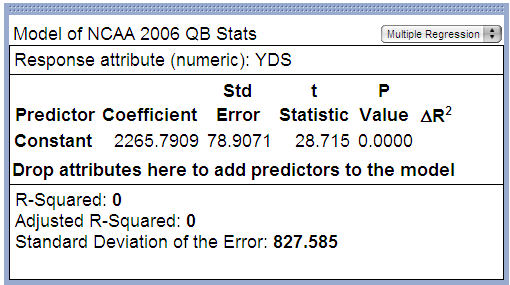
* Create a new scatterplot of ATT vs. YDS



**Finding the equation of the LSRL**

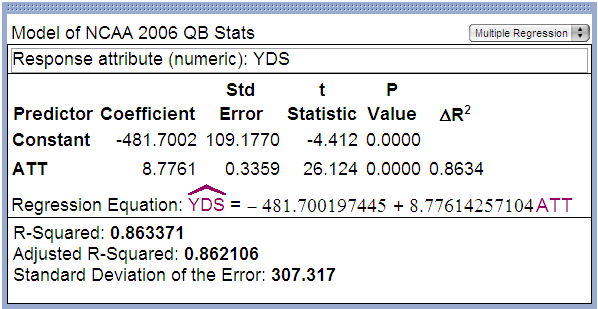
* Drag down a MODEL
* Change the type (in the top right corner) to “Multiple Regression”
* Click on the “MODEL” menu at the top of the page
* Click on “Hide Sequential Contributions chart”
* Click on the “MODEL” menu again, and then click on “Hide ANOVA table”
* Drag the Y-variable to the top, where it says

“Response attribute (numeric): unassigned”

* Drag the X-Variable to the top middle, where it says

“Drop attributes here to add predictors to the model”

* You will then see a lot of data analysis like this:



* LSR line- round the slope and Y-Intercept to 3 decimal places

Example from above:  = -481.700 + 8.776(ATT)

**Using the “Snipping tool” to copy things from Fathom**

* Click on the WINDOWS icon (bottom left corner of your computer screen
* Click on “All Programs 🡪”
* Find the Snipping tool, and click on it so it opens
* You will see the screen go gray, and then your mouse will turn into a “**+**”
* Use the mouse to “highlight” whatever it is you want to copy.
* You will then see the image you highlighted. Copy this (CTRL + C) and then go to your document (Word or Power Point) and paste the image (CTRL + V).

**PRACTICE:**

1. Open a new Fathom document.
2. Drag down a new collection.
3. Drag down a table, and label the variables X and Y
4. Enter the values from the table below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 1 | 2 | 4 | 5 | 6 | 7 | 2 | 4 | 3 | 9 | 10 | 7 |
| y | 2 | 4 | 3 | 8 | 10 | 9 | 5 | 7 | 9 | 15 | 11 | 10 |

Complete the following using the data above. Copy all things into a word document as you do them.

1. Create a scatterplot and find the correlation.
2. Now create the LSR line and find r2. Also, add the residual plot. Make the graph bigger.
3. Create the list RESID using the method above. Re-create the residual plot using this list
4. Check to see if the residuals are normally distributed. Use a normal quantile plot.
5. Create the Linear Regression output (like described above)
6. Perform a LinReg t- test to see if there is an association between X and Y.
7. Now add a new variable to your table, called GENDER. Make the first half of the entries female (just type F into the column in the table) and the second half of the entries male (just type M into the table).
8. Make a NEW scatterplot with X vs. Y. Add the LSRL. Add the residual plot. Now drag the attribute GENDER into center of that plot. This should give you a scatterplot where there are different symbols for the different genders, and there are 2 lines (one for M and one for F). Click on the male LSR line- you will see the male residual plot below. Now click on the female line- this will show you the female residual plot below. Paste both onto your work doc.
9. Save the word document as Practice 1\_last name (Example: Practice1\_McNelis)